

Coloring Random Non-Uniform Bipartite Hypergraphs

Debarghya Ghoshdastidar, Ambedkar Dukkipati
 Department of Computer Science & Automation
 Indian Institute of Science
 email: {debarghya.g,ad}@csa.iisc.ernet.in

November 18, 2015

Abstract

Let $H_{n,(p_m)_{m=2,\dots,M}}$ be a random non-uniform hypergraph of dimension M on $2n$ vertices, where the vertices are split into two disjoint sets of size n , and colored by two distinct colors. Each non-monochromatic edge of size $m = 2, \dots, M$ is independently added with probability p_m . We show that if p_2, \dots, p_M are such that the expected number of edges in the hypergraph is at least $dn \ln n$, for some $d > 0$ sufficiently large, then with probability $(1 - o(1))$, one can find a proper 2-coloring of $H_{n,(p_m)_{m=2,\dots,M}}$ in polynomial time. We present a polynomial time algorithm for hypergraph 2-coloring, and provide discussions on extension of the approach for k -coloring of non-uniform hypergraphs.

1 Introduction

A hypergraph $H = (V, E)$ is said to be bipartite or 2-colorable if the vertex set V can be partitioned into two disjoint sets V_1 and V_2 such that every edge $e \in E$ has non-empty intersections with both the partitions. In the case of graphs, one can easily find the two partitions from any given instance of H by breadth first search. However, the problem turns out to be notoriously hard if edges of size more than 2 are present. In fact, in the case of bipartite 3-uniform and 4-uniform hypergraphs, it is well known that the problem is NP-hard [11, 15].

In general, finding a proper 2-coloring is relatively easy if the hypergraph is sparse. In an answer to a question asked by Erdős [12] on 2-colorability of uniform hypergraphs, it is now known that for large m , any m -uniform hypergraph on n vertices with at most $2^{m0.7\sqrt{\frac{m}{\ln m}}}$ edges is 2-colorable [22]. As pointed in [22], the result can also be extended to non-uniform hypergraphs with minimum edge size m . However, it is much worse if the restriction on the minimum edge size and the number of hyperedges is not imposed.

Even when a hypergraph is 2-colorable, the best known algorithms [5, 8] require $O((n \ln n)^{1-1/M})$ colors to properly color the hypergraph in polynomial time, where M is the maximum edge size, also called dimension, of the hypergraph. In recent years, 2-colorability of random hypergraphs has also received considerable attention. Through a series of works [1, 9, 21], it is now established that random uniform hypergraphs are 2-colorable only when the number of edges are at most Cn , for some constant $C > 0$. Thus, it is evident that coloring relatively dense hypergraphs is difficult unless the hypergraph admits a “nice” structure.

In spite of the hardness of the problem, there are a number of applications that require hypergraph coloring algorithms. For instance, such algorithms have been used for approximate DNF counting [18], as well as in various resource allocation and scheduling problems [7, 3]. The connection between “Not-All-Equal” (NAE) SAT and hypergraph 2-coloring also demonstrate its significance in context of satisfiability problems. Among the various approaches studied in the literature, perhaps the only known non-probabilistic instances of efficient 2-coloring are in the cases where the hypergraph is α -dense, 3-uniform and bipartite [8], or where the hypergraph is m -uniform and its every edge has equal number of vertices of either colors [19].

In this paper, we consider the problem of coloring random non-uniform hypergraphs of dimension M , that has an underlying planted bipartite structure. We present a polynomial time algorithm that can properly 2-color instances of the random hypergraph with high probability whenever the expected number of edges is at least $dn \ln n$ for some constant $d > 0$. To the best of our knowledge, such a model has been only considered by Chen and Frieze [8], who extended a graph coloring approach of Alon and Kahale [4] to present an algorithm for 2-coloring of 3-uniform bipartite hypergraphs with dn number of edges. To this end, our work generalizes the results of [8] to non-uniform hypergraphs, and it is the first algorithm that is guaranteed to properly color non-uniform bipartite hypergraphs using only two colors. We also discuss the possible extension of our approach to the case of non-uniform k -colorable hypergraphs.

The Main Result

Before stating the main result of this paper, we present the planted model under consideration, which is based on the model that is studied in [14]. The random hypergraph $H_{n, (p_m)_{m=2, \dots, M}}$ is generated on the set of vertices $V = \{1, 2, \dots, 2n\}$, which is arbitrarily split into two sets, each of size n , and the sets are colored with two different colors. Given an integer M , and $p_2, \dots, p_M \in [0, 1]$, the edges of the hypergraph are randomly added in the following way. All the edges of size at most M are added independently,

and for any $e \subset V$,

$$P(e \in E) = \begin{cases} p_m & \text{if } e \text{ is not monochromatic and } |e| = m, \\ 0 & \text{otherwise.} \end{cases}$$

We prove the following result.

Theorem 1. *Assume $M = O(1)$. There is a constant $d > 0$ such that if*

$$\sum_{m=2}^M p_m \binom{2n}{m} \geq dn \ln n, \quad (1)$$

then with probability $(1 - o(1))$, Algorithm COLOR (presented in next section) finds a proper 2-coloring of the random non-uniform bipartite hypergraph $H_{n, (p_m)_{m=2, \dots, M}}$.

It is easy to see that the expected number of edges in the hypergraph is $\Theta\left(\sum_{m=2}^M p_m \binom{2n}{m}\right)$, and so the condition may be stated in terms of expected number of edges.

Organization of this paper

The rest of the paper is organized in the following manner. In Section 2, we present our coloring algorithm, followed by a proof of Theorem 1 in Section 3. In the concluding remarks in Section 4, we provide discussions about the key assumptions made in this work, and also the possible extensions of our results to k -coloring and strong coloring of non-uniform hypergraphs. The appendix contains proofs of the lemmas mentioned in Section 3.

2 Spectral algorithm for hypergraph coloring

The coloring algorithm, presented below, is similar in spirit to the spectral methods of [4, 8], but certain key differences exist, which are essential to deal with non-uniform hypergraphs.

Given a hypergraph $H = (V, E)$, an initial guess of the color classes is formed by exploiting the spectral properties of a certain matrix $A \in \mathbb{R}^{|V| \times |V|}$ defined as

$$A_{ij} = \begin{cases} \sum_{e \in E: e \ni i, j} \frac{1}{|e|} & \text{if } i \neq j, \text{ and} \\ \sum_{e \in E: e \ni i} \frac{1}{|e|} & \text{if } i = j. \end{cases} \quad (2)$$

The above matrix has been used in the literature to construct the Laplacian of a hypergraph [6, 14], and is also known to be related to the affinity matrix

of the star expansion of hypergraph [2]. The use of matrix A is in contrast to the adjacency based graph construction of [8] that is likely to result in a complete graph if the hypergraph is dense.

The later stage of the algorithm considers an iterative procedure that is similar to [4, 8], but uses a weighted summation of neighbors. Such weighting is crucial while dealing with the edges of different sizes.

Algorithm COLOR – Colors a non-uniform hypergraph H :

- 1: Define the matrix A as in (2).
 - 2: Compute $x^A = \arg \min_{\|x\|_2=1} x^T A x$.
 - 3: Let $T = \lceil \log_2 n \rceil$, $V_1^{(0)} = \{i \in V : x_i^A \geq 0\}$ and $V_2^{(0)} = \{i \in V : x_i^A < 0\}$.
 - 4: **for** $t = 1, 2, \dots, T$ **do**
 - 5: Let $V_1^{(t)} = \left\{ i \in V : \sum_{j \in V_1^{(t-1)} \setminus \{i\}} A_{ij} < \sum_{j \in V_2^{(t-1)} \setminus \{i\}} A_{ij} \right\}$,
 and $V_2^{(t)} = V \setminus V_1^{(t)}$.
 - 6: **end for**
 - 7: **if** $\exists e \in E$ such that $e \subset V_1^{(T)}$ or $e \subset V_2^{(T)}$ **then**
 - 8: Algorithm FAILS.
 - 9: **else**
 - 10: 2-Color V according to the partitions $V_1^{(T)}, V_2^{(T)}$.
 - 11: **end if**
-

3 Proof of Main Result

We now prove Theorem 1. Without loss of generality, assume that the true color classes in V are $\{1, 2, \dots, n\}$ and $\{n+1, \dots, 2n\}$. Also, let $W^{(t)}$, $t = 0, 1, \dots, T$, denote the incorrectly colored vertices after iteration t , with $W^{(0)}$ being the incorrectly colored nodes after initial spectral step. We prove Theorem 1 by showing with probability $(1 - o(1))$, the size of $W^{(T)} < 1$, which implies that all nodes are correctly colored, and hence, the hypergraph must be properly colored.

The first lemma bounds the size of $W^{(0)}$, *i.e.*, the error incurred at the initial spectral step.

Lemma 1. *With probability $(1 - o(1))$, $|W^{(0)}| \leq \frac{n}{M^2 2^{2M+4}}$.*

Next, we analyze the iterative stage of the algorithm to make the following claim, which characterizes the vertices that are correctly colored after iteration t .

Lemma 2. Let $\eta = \frac{1}{2^{M+2}} \sum_{m=2}^M \frac{p_m(n-1)}{m} \binom{n-2}{m-2}$. For any $t \in \{1, \dots, T\}$, if $\sum_{j \in W^{(t-1)} \setminus \{i\}} A_{ij} < \eta$ for any $i \in V$, then $P(i \in W^{(t)}) \leq n^{-\Omega(d)}$.

Note that there are only $T = \lceil \log_2 n \rceil$ iterations, and $|V| = 2n$. Combining the result of Lemma 2 with union bound, we can conclude that with probability $(1 - o(1))$, for all iterations $t = 1, 2, \dots, T$, there does not exist any $i \in V$ such that $\sum_{j \in W^{(t-1)} \setminus \{i\}} A_{ij} < \eta$. We also make the following observation, where η is defined in Lemma 2.

Lemma 3. With probability $(1 - o(1))$, there does not exist $C_1, C_2 \subset V$ such that $|C_1| \leq \frac{n}{M^2 2^{2M+4}}$, $|C_2| = \frac{1}{2}|C_1|$ and for all $i \in C_2$, $\sum_{j \in C_1 \setminus \{i\}} A_{ij} \geq \eta$.

We now use the above lemmas to proceed with the proof of Theorem 1. Lemma 1 shows that $|W^{(0)}| \leq \frac{n}{M^2 2^{2M+4}}$ with probability $(1 - o(1))$. Conditioned on this event, and due to the conclusion of Lemma 2, one can argue that Lemma 3 is violated unless $|W^{(t)}| < \frac{1}{2}|W^{(t-1)}|$ for all iteration t with probability $(1 - o(1))$. Thus, in each iteration, the number of incorrectly colored vertices are reduced by at least half. Hence, after $T = \lceil \log_2 n \rceil$ iterations, $|W^{(T)}| < 1$, which implies that all vertices are correctly colored.

4 Discussions and Concluding remarks

In this paper, we showed that a random non-uniform bipartite hypergraph of dimension M with balanced partitions can be properly 2-colored with probability $(1 - o(1))$ by a polynomial time algorithm. The proposed method uses a spectral approach to form initial guess of the color classes, which is further refined iteratively. To the best of our knowledge, this is the first work on 2-coloring bipartite non-uniform hypergraphs. Previous works [8, 16] have only restricted to the case of uniform hypergraphs.

A note on the assumptions in Theorem 1

The key assumptions made in this paper are the following:

1. $M = O(1)$, and
2. p_2, \dots, p_M are such that the expected number of edges is larger than $dn \ln n$, where $d > 0$ is a large constant.

The assumption $M = O(1)$ is crucial, particularly in Lemma 1, and helps to ensure that d can be chosen to be a constant. This can be avoided if d is allowed to increase with n appropriately. We note that a previous work on

spectral hypergraph partitioning [14] allows M to grow with n , but imposes an additional restriction so that the number of edges of larger size decay rapidly.

The second assumption is stronger than the one in [8], where it was shown that a random bipartite 3-uniform hypergraph can be properly 2-colored with high probability if the expected number of edges is dn . This is due to the use of matrix Bernstein inequality [23] in Lemma 1 that does not provide useful bounds in the most sparse case. On the other hand, Chen and Frieze [8] use the techniques of Kahn and Szemerédi [13] that allows them to work in the most sparse regime. However, it is not clear how the same techniques can be extended even to uniform hypergraphs of higher order. Thus, it remains an open problem whether a similar result can be proved when the number of edges in the hypergraph grows linearly with n .

k -coloring of hypergraphs

Though Algorithm COLOR has been presented only for the hypergraph 2-coloring problem, one may easily extend the approach to achieve a k -coloring, where the objective is to color the vertices of the hypergraph with k colors such that no edge is monochromatic. A possible extension of Algorithm COLOR is as follows:

1. In Step 2, compute the eigenvectors corresponding to the $(k-1)$ smallest eigenvalues of A .
2. Use k -means algorithm [20] to cluster rows of the eigenvector matrix into k groups, and define the initial guess for the color classes $V_1^{(0)}, \dots, V_k^{(0)}$ in Step 3 according to the above clustering.
3. The iterative computation in Step 6 is modified by defining

$$V_l^{(t)} = \left\{ i \in V : \sum_{j \in V_l^{(t-1)} \setminus \{i\}} A_{ij} < \sum_{j \in V_{l'}^{(t-1)} \setminus \{i\}} A_{ij} \text{ for all } l' \neq l \right\}$$

for $l = 1, 2, \dots, (k-1)$, and $V_k^{(t)} = V \setminus \left(\bigcup_{l < k} V_l^{(t)} \right)$.

In the above modification, we borrow the popular idea of using k -means on the rows of eigenvector matrix to find k planted partitions in a graph or hypergraph [17, 14].

We believe that the result in Theorem 1 can be extended to this setting, where the random model allows for k planted color classes in the hypergraph with non-monochromatic edges generated in the aforementioned manner. Assuming $k = O(1)$ and k -means algorithm always provides a near optimal solution, one can follow the arguments of [14] to prove a result similar to

Lemma 1. On the other hand, Lemmas 2 and 3 should hold for an appropriate choice of η . Hence, one can comment that the algorithm achieves a proper k -coloring with probability $(1 - o(1))$.

We also note that Algorithm COLOR can be used for finding solutions of NAE-SAT problems. The extension of COLOR is also applicable for strong coloring of hypergraphs, which finds applications in design of communication networks [24].

Proofs of technical lemmas

Proof of Lemma 1

We view the random matrix $A \in \mathbb{R}^{2n \times 2n}$, as a perturbation of its expected value $\mathcal{A} = \mathbb{E}[A]$. Let \mathcal{E} denote the collection of all the non-monochromatic subsets of V of size at most M . One can verify that for any $i, j \in V$, $i \neq j$

$$\mathcal{A}_{ij} = \sum_{e \in \mathcal{E}: e \ni i, j} \frac{p|e|}{|e|} \quad \text{and} \quad \mathcal{A}_{ii} = \sum_{e \in \mathcal{E}: e \ni i} \frac{p|e|}{|e|}.$$

Counting the number of possible edges of each size, one can see that

$$\mathcal{A}_{ij} = \begin{cases} \alpha_1 - \alpha_2 & \text{if } i \neq j, \text{ and } i, j \text{ belong to same color class,} \\ \alpha_1 & \text{if } i \neq j, \text{ and } i, j \text{ belong to different color class,} \\ \alpha_1 - \alpha_2 + \alpha_3 & \text{if } i = j, \end{cases} \quad (3)$$

where

$$\begin{aligned} \alpha_1 &= \sum_{m=2}^M \frac{p_m}{m} \binom{2n-2}{m-2}, & \alpha_2 &= \sum_{m=2}^M \frac{p_m}{m} \binom{n-2}{m-2}, \\ \text{and } \alpha_3 &= \sum_{m=2}^M \frac{p_m}{m} \left(\binom{2n-2}{m-1} - \binom{n-2}{m-1} \right). \end{aligned}$$

Hence, we can write \mathcal{A} as

$$\mathcal{A} = \alpha_1 1_{2n \times 2n} - \alpha_2 \begin{pmatrix} 1_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & 1_{n \times n} \end{pmatrix} + \alpha_3 I_{2n}, \quad (4)$$

where I_{2n} is the $2n$ -dimensional identity matrix, and $1_{n \times n}$ is a $n \times n$ matrix of all 1's. One can verify that the smallest eigenvalue of \mathcal{A} is $(\alpha_3 - n\alpha_2)$, which has multiplicity 1, and is separated from the other eigenvalues by an eigen-gap of $n\alpha_2$. Moreover, the corresponding unit norm eigenvector $x^{\mathcal{A}}$ is such that $x_i^{\mathcal{A}} = \frac{1}{\sqrt{2n}}$ for all $i \leq n$, and $x_i^{\mathcal{A}} = -\frac{1}{\sqrt{2n}}$ for all $i > n$, up to a possible change of sign.

At this stage, we refer to a well-known result from matrix perturbation theory [10]. We state the result in a particular form that is appropriate in our setting. The result, as stated in Theorem 2, has been previously used in [14, Lemma 4.4] and [17].

Theorem 2 (Davis-Kahan $\sin \Theta$ theorem). *Let $\mathcal{A} \in \mathbb{R}^{d \times d}$ be a symmetric matrix, and A be an additive perturbation of \mathcal{A} . Let $S \subset \mathbb{R}$ be any interval that contains exactly k eigenvalues of \mathcal{A} . Define*

$$\delta = \min\{|\lambda - \lambda'| : \lambda \in S, \lambda' \notin S, \text{ and } \lambda, \lambda' \text{ are eigenvalues of } \mathcal{A}\}.$$

If $\delta > 2\|A - \mathcal{A}\|_2$, then S also contains exactly k eigenvalues of A .

Let $X, \mathcal{X} \in \mathbb{R}^{d \times k}$ be orthonormal eigenvector matrices for the eigenvalues in S of A, \mathcal{A} respectively. Then there is an orthonormal (rotation) matrix $Q \in \mathbb{R}^{k \times k}$ such that

$$\|X - \mathcal{X}Q\|_F \leq \frac{2\sqrt{2k}\|A - \mathcal{A}\|_2}{\delta}.$$

By viewing A as a perturbation of \mathcal{A} and noting that the eigen-gap $\delta = n\alpha_2$, one can use Theorem 2 to conclude that if $\alpha_2 > \frac{2}{n}\|A - \mathcal{A}\|_2$, then

$$\|x^A - x^{\mathcal{A}}\|_2 \leq \frac{2\sqrt{2}\|A - \mathcal{A}\|_2}{n\alpha_2}. \quad (5)$$

One can write A as $A = \sum_{e \in \mathcal{E}} \frac{h_e}{|e|} a_e a_e^T$, where, for each set $e \in \mathcal{E}$, h_e is a Bernoulli($p_{|e|}$) random variable, and $a_e \in \{0, 1\}^{2n}$ is such that $(a_e)_i = 1$ only when $i \in e$. Hence, one may view A as a sum of independent random matrices. To this end, the following concentration inequality is quite useful to derive a bound on the perturbation $\|A - \mathcal{A}\|_2$.

Theorem 3 (Matrix Bernstein inequality [23]). *Consider a finite sequence X_1, X_2, \dots, X_L of independent, random, self-adjoint matrices with dimension d . Assume that each random matrix satisfies $\|X_l - \mathbb{E}[X_l]\|_2 \leq R$ almost surely. Define $X = \sum_{l=1}^L X_l$, and let $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$, where we assume all the above expectations exist. Then for all $t > 0$,*

$$\mathbb{P}(\|X - \mathbb{E}[X]\|_2 \geq t) \leq d \exp\left(\frac{-t^2}{2\text{Var}(X) + \frac{2}{3}Rt}\right).$$

The above result directly implies

$$\mathbb{P}(\|A - \mathcal{A}\|_2 > 4\sqrt{n\alpha_1 \ln n}) \leq 4n \exp\left(-\frac{16n\alpha_1 \ln n}{2\|\text{Var}(A)\|_2 + \frac{8}{3}\sqrt{n\alpha_1 \ln n}}\right). \quad (6)$$

We note that choosing d large enough, one can satisfy $n\alpha_1 > \ln n$. Also, observe that

$$\|\text{Var}(A)\|_2 \leq \max_i \sum_{j=1}^{2n} (\text{Var}(A))_{ij} \leq \max_i \sum_{j=1}^{2n} \mathcal{A}_{ij} \leq 4n\alpha_1.$$

Substituting these in (6), we have

$$\begin{aligned} \mathbb{P}(\|A - \mathcal{A}\|_2 > 4\sqrt{n\alpha_1 \ln n}) &\leq 4n \exp\left(-\frac{16n\alpha_1 \ln n}{8n\alpha_1 + \frac{8}{3}n\alpha_1}\right) \\ &= \frac{4}{\sqrt{n}} = o(1). \end{aligned} \quad (7)$$

Thus, with probability $(1 - o(1))$ we have $\|A - \mathcal{A}\|_2 \leq 4\sqrt{n\alpha_1 \ln n}$. Due to this bound, one can argue that if $n\alpha_2 > 8\sqrt{\alpha_1 n \ln n}$, i.e., $\frac{\alpha_1}{\alpha_2} < \frac{n}{64 \ln n}$, then the condition in Theorem 2 is satisfied, and the perturbation bound (5) holds. We can compute that

$$\begin{aligned} \frac{\alpha_1}{\alpha_2} &= \frac{\sum_{m=2}^M \frac{p_m}{m} \binom{2n-2}{m-2}}{\left(\sum_{m=2}^M \frac{p_m}{m} \binom{n-2}{m-2}\right)^2} \\ &\leq \frac{n^2 2^{2M+2}}{\sum_{m=2}^M p_m (m-1) \binom{2n}{m}} \\ &\leq \frac{n 2^{2M+2}}{d \ln n}. \end{aligned}$$

Hence, choosing d sufficiently large, the above mentioned condition holds, and one can claim from (5) that

$$\|x^A - x^{\mathcal{A}}\|_2 \leq \frac{8\sqrt{2n\alpha_1 \ln n}}{n\alpha_2} \leq \frac{2^{M+4.5}}{\sqrt{d}}.$$

Now, we define the set $\widehat{W} \subset V$ as $\widehat{W} = \{i \in V : |x_i^A - x_i^{\mathcal{A}}| \geq \frac{1}{\sqrt{2n}}\}$. From the definition of the color classes $V_1^{(0)}, V_2^{(0)}$, it directly follows that any vertex not in \widehat{W} must be correctly colored. Hence,

$$\begin{aligned} |W^{(0)}| &\leq |\widehat{W}| \\ &\leq \sum_{i \in \widehat{W}} 2n |x_i^A - x_i^{\mathcal{A}}|^2 \\ &\leq 2n \|x^A - x^{\mathcal{A}}\|_2^2 \\ &= O\left(\frac{n}{d}\right), \end{aligned}$$

where the bound holds with probability $(1 - o(1))$. Thus, choosing d sufficiently large, one obtains that $|W^{(0)}| \leq \frac{n}{M^2 2^{2M+4}}$.

Proof of Lemma 2

Consider any $i \leq n$. Note that i is correctly colored in iteration t if

$$\sum_{j \in V_1^{(t-1)} \setminus \{i\}} A_{ij} < \sum_{j \in V_2^{(t-1)} \setminus \{i\}} A_{ij},$$

or equivalently,

$$\sum_{j \in V_1^{(t-1)} \setminus \{i\}} A_{ij} < \frac{1}{2} \sum_{j \neq i} A_{ij}. \quad (8)$$

Hence, it suffices to show that (8) holds under the condition stated in the lemma. A similar condition can be stated for $i > n$.

We note that $\sum_{j \neq i} A_{ij} = \sum_{e \in \mathcal{E}: e \ni i} h_e \frac{(|e| - 1)}{|e|}$, and so, from Bernstein inequality, we have

$$\begin{aligned} & \mathbb{P} \left(\sum_{j \neq i} A_{ij} \leq \left(1 - \frac{1}{2^{M+2}} \right) \sum_{j \neq i} \mathcal{A}_{ij} \right) \\ & \leq \exp \left(- \frac{\frac{1}{2^{2M+4}} \left(\sum_{j \neq i} \mathcal{A}_{ij} \right)^2}{2 \sum_{e \in \mathcal{E}: e \ni i} \frac{(|e|-1)^2}{|e|^2} \text{Var}(h_e) + \frac{2}{3 \cdot 2^{M+2}} \sum_{j \neq i} \mathcal{A}_{ij}} \right) \\ & \leq \exp \left(- \Omega \left(\sum_{j \neq i} \mathcal{A}_{ij} \right) \right) \\ & \leq n^{-\Omega(d)}. \end{aligned}$$

The second inequality holds since for any e , $\frac{(|e|-1)^2}{|e|^2} \text{Var}(h_e) \leq \frac{(|e|-1)}{|e|} \mathbb{E} h_e$, and the last inequality is true under the condition of Theorem 1 since

$$\begin{aligned} \sum_{j \neq i} \mathcal{A}_{ij} &= (2n - 1)\alpha_1 + (n - 1)\alpha_2 \\ &= \sum_{m=2}^M \frac{p_m(m-1)}{2n} \left[\binom{2n}{m} - 2 \binom{n}{m} \right] \\ &= \Omega(d \ln n). \end{aligned}$$

Denoting $[n - i] = \{1, \dots, n\} \setminus i$, *i.e.*, the first color class excluding vertex i ,

we have $\sum_{j \in [n-i]} A_{ij} = \sum_{e \in \mathcal{E}: e \ni i} h_e \frac{|e \cap [n-i]|}{|e|}$, and one can bound

$$\begin{aligned} & \mathbb{P} \left(\sum_{j \in [n-i]} A_{ij} \geq \left(1 + \frac{1}{2^{M+2}} \right) \sum_{j \in [n-i]} \mathcal{A}_{ij} \right) \\ & \leq \exp \left(- \frac{\frac{1}{2^{2M+4}} \left(\sum_{j \in [n-i]} \mathcal{A}_{ij} \right)^2}{2 \sum_{e \in \mathcal{E}: e \ni i} \text{Var}(h_e) \frac{|e \cap U|^2}{|e|^2} + \frac{2}{3 \cdot 2^{M+2}} \sum_{j \in [n-i]} \mathcal{A}_{ij}} \right) \\ & \leq n^{-\Omega(d)}. \end{aligned}$$

Thus, with probability $(1 - n^{-\Omega(d)})$, we have

$$\begin{aligned} \sum_{j \in [n-i]} A_{ij} & < \left(1 + \frac{1}{2^{M+2}} \right) \sum_{j \in [n-i]} \mathcal{A}_{ij} \\ & = \sum_{m=2}^M \frac{p_m(n-1)}{m} \left(1 + \frac{1}{2^{M+2}} \right) \left(\binom{2n-2}{m-2} - \binom{n-2}{m-2} \right), \end{aligned}$$

and

$$\begin{aligned} \sum_{j \neq i} A_{ij} & > \left(1 - \frac{1}{2^{M+2}} \right) \sum_{j \neq i} \mathcal{A}_{ij} \\ & = \sum_{m=2}^M \frac{p_m}{m} \left(1 - \frac{1}{2^{M+2}} \right) \left((2n-1) \binom{2n-2}{m-2} - (n-1) \binom{n-2}{m-2} \right). \end{aligned}$$

Using above relation, we can derive (8) since

$$\begin{aligned} \sum_{j \in V_1^{(t-1)} \setminus \{i\}} A_{ij} & = \sum_{j \in W^{(t-1)} \cap V_1^{(t-1)} \setminus \{i\}} A_{ij} + \sum_{j \in V_1^{(t-1)} \setminus (W^{(t-1)} \cap \{i\})} A_{ij} \\ & \leq \sum_{j \in W^{(t-1)} \setminus \{i\}} A_{ij} + \sum_{j \in [n-i]} A_{ij} \\ & < \eta + \left(1 + \frac{1}{2^{M+2}} \right) \sum_{j \in [n-i]} \mathcal{A}_{ij} \end{aligned}$$

The first inequality uses the fact $V_1^{(t-1)} \setminus W^{(t-1)}$ is the set of correctly colored nodes, with true color same as i . Hence, $V_1^{(t-1)} \setminus (W^{(t-1)} \cap \{i\}) \subset [n-i]$.

From definition of η , we have

$$\begin{aligned}
& \sum_{j \in V_1^{(t-1)} \setminus \{i\}} A_{ij} \\
& \leq \sum_{m=2}^M \frac{p_m(n-1)}{m} \left[\frac{1}{2^{M+2}} \binom{n-2}{m-2} + \left(1 + \frac{1}{2^{M+2}} \right) \left(\binom{2n-2}{m-2} - \binom{n-2}{m-2} \right) \right] \\
& = \sum_{m=2}^M \frac{p_m(n-1)}{m} \left(1 - \frac{1}{2^{M+2}} \right) \left[\binom{2n-2}{m-2} - \frac{1}{2} \binom{n-2}{m-2} \right] \\
& + \sum_{m=2}^M \frac{p_m(n-1)}{2m} \left[\frac{1}{2^M} \binom{2n-2}{m-2} - \binom{n-2}{m-2} \right] - \sum_{m=2}^M \frac{p_m(n-1)}{m 2^{M+3}} \binom{n-2}{m-2}.
\end{aligned}$$

One can see that the first term is at most $\frac{1}{2} \left(1 - \frac{1}{2^{M+2}} \right) \sum_{j \neq i} \mathcal{A}_{ij} < \frac{1}{2} \sum_{j \neq i} A_{ij}$. On the other hand, we note that

$$\frac{\binom{2n-2}{m-2}}{\binom{n-2}{m-2}} \leq \frac{1}{4} \frac{\binom{2n}{m}}{\binom{n}{m}} \leq \frac{1}{4} \frac{(2n)^m}{\frac{n^m}{4 \cdot m!}} = 2^m \leq 2^M.$$

So the second term is negative, which proves (8), and the claim follows.

Proof of Lemma 3

Let $C_1, C_2 \subset V$ be arbitrary such that $|C_2| = b$, and $E_{C_1 C_2}$ be the set of all non-monochromatic subsets of V of size at most M that have non-empty intersection with both C_1 and C_2 . Then

$$\begin{aligned}
\sum_{e \in E_{C_1 C_2}} h_e & \geq \frac{1}{M} \sum_{e \in E_{C_1 C_2}} h_e \frac{|e \cap C_1| |e \cap C_2|}{|e|} \\
& \geq \frac{1}{M} \sum_{i \in C_2} \sum_{j \in C_1 \setminus \{i\}} A_{ij} \geq \frac{b\eta}{M},
\end{aligned}$$

where the last inequality holds under the condition stated in the lemma. Now we bound the probability

$$\begin{aligned}
& \mathbb{P} \left(\exists C_1, C_2 \subset V, |C_2| = \frac{1}{2}|C_1| \leq \frac{n}{M^2 2^{2M+5}}, \sum_{j \in C_1 \setminus \{i\}} A_{ij} \geq \eta \ \forall i \in C_2 \right) \quad (9) \\
& \leq \sum_{b=1}^{\frac{n}{M^2 2^{2M+5}}} \mathbb{P} \left(\exists C_1, C_2 \subset V, |C_2| = \frac{1}{2}|C_1| = b, \text{ and } \sum_{e \in E_{C_1 C_2}} h_e \geq \frac{b\eta}{M} \right) \\
& \leq \sum_{b=1}^{\frac{n}{M^2 2^{2M+5}}} \sum_{C_2: |C_2|=b} \sum_{C_1: |C_1|=2b} \mathbb{P} \left(\sum_{e \in E_{C_1 C_2}} h_e \geq \frac{b\eta}{M} \right)
\end{aligned}$$

We observe that

$$\begin{aligned}
\sum_{e \in E_{C_1 C_2}} \mathbb{E}[h_e] &= \sum_{m=2}^M \sum_{e \in E_{C_1 C_2}, |e|=m} p_m \\
&\leq 2b^2 \sum_{m=2}^M p_m \binom{2n-2}{m-2} \\
&\leq b^2 2^{M+1} \sum_{m=2}^M p_m \binom{n-2}{m-2} \\
&\leq \frac{b^2 \eta M 2^{2M+4}}{n},
\end{aligned}$$

and the above bound is smaller than $\frac{b\eta}{2M}$ for $b \leq \frac{n}{M^2 2^{2M+5}}$. Hence, we can write

$$\begin{aligned}
&\mathbb{P} \left(\sum_{e \in E_{C_1 C_2}} h_e \geq \frac{b\eta}{M} \right) \\
&\leq \exp \left(\frac{- \left(\frac{b\eta}{M} - \sum_{e \in E_{C_1 C_2}} \mathbb{E}[h_e] \right)^2}{2 \sum_{e \in E_{C_1 C_2}} \text{Var}(h_e) + \frac{2}{3} \left(\frac{b\eta}{M} - \sum_{e \in E_{C_1 C_2}} \mathbb{E}[h_e] \right)} \right) \\
&\leq \exp \left(-\frac{3b\eta}{16M} \right).
\end{aligned}$$

Substituting in (9), we have the probability of the existence of C_1, C_2 with mentioned conditions is at most

$$\sum_{b=1}^{\frac{n}{M^2 2^{2M+5}}} \binom{2n}{b} \binom{2n}{2b} \exp \left(-\frac{3b\eta}{16M} \right) \leq \sum_{b=1}^{\infty} \left(2n \exp \left(1 - \frac{\eta}{16M} \right) \right)^{3b}.$$

Under the assumption of Theorem 1, one can verify that $\eta \geq \frac{d \ln n}{2^{2M+4}}$. So for large d , the above geometric series converges, and is at most $n^{-\Omega(d)} = o(1)$. Hence, the claim.

References

- [1] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *Proceedings of 49th Annual Symposium on Foundations of Computer Science*, 2008.
- [2] S. Agarwal, K. Branson, and S. Belongie. Higher order learning with graphs. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 17–24, 2006.

- [3] N. Ahuja and A. Srivastava. On constrained hypergraph coloring and scheduling. In *Approximation Algorithms for Combinatorial Optimization*, pages 14–25. Springer Berlin Heidelberg, 2002.
- [4] N. Alon and N. Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal of Computing*, 26:1733–1748, 1997.
- [5] N. Alon, P. Kelsen, S. Mahajan, and R. Hariharan. Coloring 2-colorable hypergraphs with a sublinear number of colors. *Nordic Journal of Computing*, 3:425–439, 1996.
- [6] M. Bolla. Spectra, euclidean representations and clusterings of hypergraphs. *Discrete Mathematics*, 117(1):19–39, 1993.
- [7] A. Capitanio, A. Nicolau, and N. Dutt. A hypergraph-based model for port allocation on multiple-register-file vliw architectures. *International Journal of Parallel Programming*, 23(6):499–513, 1995.
- [8] H. Chen and A. Frieze. Coloring bipartite hypergraphs. In *Integer Programming and Combinatorial Optimization*, pages 345–358, 1996.
- [9] A. Coja-Oghlan and L. Zdeborová. The condensation transition in random hypergraph 2-coloring. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2012.
- [10] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- [11] I. Dinur, O. Regev, and C. D. Smyth. The hardness of 3-uniform hypergraph coloring. *Combinatorica*, 25(1):519–535, 2005.
- [12] P. Erdős. On a combinatorial problem. *Nordisk Mat. Tidskr*, 11:5–10, 1963.
- [13] J. Friedman, J. Kahn, and E. Szemerédi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM Symposium on Theory of Computing*, 1989.
- [14] D. Ghoshdastidar and A. Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *arXiv preprint*, 2015.
- [15] S. Khot and R. Saket. Hardness of finding independent sets in 2-colorable and almost 2-colorable hypergraphs. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1607–1625, 2014.
- [16] M. Krivelevich and B. Sudakov. Approximate coloring of uniform hypergraphs. *Journal of Algorithms*, 49:2–12, 2003.

- [17] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, 2015.
- [18] C.-J. Lu. Deterministic hypergraph coloring and its applications. *SIAM Journal on Discrete Mathematics*, 18(2):320–331, 2004.
- [19] C. McDiarmid. A random recolouring method for graphs and hypergraphs. *Combinatorics, Probability and Computing*, 2(3):215–229, 1993.
- [20] R. Ostrovsky, Y. Rabani, L. J. Schulman, and C. Swamy. The effectiveness of lloyd-type methods for the k-means problem. *Journal of the ACM*, 59(6):28:1–28:22, 2013.
- [21] K. Panagiotou and A. Coja-Oghlan. Catching the k-NAESAT threshold. In *ACM Symposium on Theory of Computing*. ACM, 2012.
- [22] J. Radhakrishnan and A. Srinivasan. Improved bounds and algorithms for hypergraph two-coloring. In *Proceedings of 39th Annual Symposium on Foundations of Computer Science*, pages 684–693, 1998.
- [23] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.
- [24] S. Wu, S. Wei, Y. Wang, R. Vaidyanathan, and J. Yuan. Partition information and its transmission over partition information and its transmission over boolean multi-access channels. *IEEE Transactions on Information Theory*, 61(2):1010–1027, 2015.